THE CHRONOLOGY OF OLD ROMANIAN WORDS

Daniela Gîfu

Scientific Researcher III, PhD, "Alexandru. Ioan Cuza" University of Iași

Abstract. The paper describes an analysis of the Romanian newspapers from the second decade of the 18th century in order to develop an old lexicon (significant differences of morphology and syntax appear). During the last year, texts from the print press, starting with 1829, were collected and written, covering three Romanian regions (Moldovia, Transylvania and Wallachia). This collection of texts is automatically processed with natural language processing tools, consisting in: segmentation, tokenization, lemmatization, POS-tagging, NP-chunking, name entities recognition (to create a gazetier with old toponyms). In this collection of articles, an important number of old words are extracted and marked with two labels (year and region) which attests their origin. Furthermore, every entry in this old vocabulary is searched in eDTLR (Dictionary Thesaurus of the Romanian Language in electronic form) in order to improve the chronological data (of course, taking into account the documents included in it). Such research comes to support direct beneficiaries (lexicographers, anthropologists, journalists, specialists in communication), and anyone concerned about this topic.

Keywords: old words, collection of Romanian newspapers, natural language processing, E-DTLR, chronology.

1. Introducere

Din varii motive, de cele mai multe ori istorice, similitudinile lingvistice între două limbi surori au ponderi diferite în anumite perioade de timp. Este aproape imposibil ca un corpus construit doar prin efort uman să acopere un demers de măsurare statistică a similitudinii lingvistice. Lucrarea de față propune o metodologie de investigare diacronică plecând de la 4 colecții de articole jurnalistice din Moldova, Transilvania, Valahia și Bucovina, începând cu a doua decadă a secolului al XVIII-lea până în prezent cu scopul de

Arhipelag XXI Press, Tîrgu Mureș, ISBN: 978-606-8624-03-7

BDD-V3969 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 00:54:25 UTC)

construi 4 resurse lexicale[1] structurate din cuvinte necunocute, etichetate cu NotInDict[2] pe baza cărora se va încerca o recuperare cronologică a acestora, raportându-ne la Dictionarul Tezaur al Limbii Române în versiune electronică.

Resursele lingvistice sunt văzute drept "izvoarele fundamentale de informații cu privire la limba română (n.n.), stocate convenabil (chiar dacă încă preponderent în maniera tradițională) și care, în calitate de componente ale culturii în sensul cel mai larg, sunt susceptibile de a fi valorificate pentru studierea limbii române" [Vintilă-Rădulescu, 2002]. Pentru limba română, cele mai reprezentative resurse lingvistice au fost realizate la Institutul de Lingvistică "Iorgu Iordan", astăzi Institutul de Lingvistică "Iorgu Iordan - Al. Rosetti" [Avram et al., 1999].

Lucrarea este structurată în cinci capitole. După o scurtă introducere cu privire la tema propusă în care întărim ideea importanței dezvoltării resurselor lexicale (dicționare, corpusuri etc.), în capitolul trei descriem pe scurt resursele lexicale care fac parte din acest studiu. În capitolul patru prezentăm metodologia de lucru centrată pe identificarea cuvintelor vechi din corpusul prezent (1829-2015), care au fost etichetate cu anul și regiunea din care au fost extrase, urmând ca pe baza eDTLR-ului să se recupereze automat cronologia acelor cuvinte în timp și spațiu pe baza citatelor. Încheiem cu concluziile prezentului studiu și pașii viitori de cercetare, axați pe analiza contrastivă dintre limbile scrise surori.

2. Context

În contextul societăţii informatizate, limba devine tehnologie, subiect al investigaţiei tehnice (Danzin, 1992; Tufiş, & Slama-Cazacu, 1994). Tehnologia limbajului impune metodologii specifice de cercetare/dezvoltare, alinierea la standardele de facto deja existente în codificarea resurselor lingvistice.

Informatizarea limbii române a devenit o prioritate pentru lingviștii informaticieni care definesc și implementează diverse modele formale în dorința dezvoltării unui sistem automat de prelucrare lingvistică semnificativă. Limba fiind un fenomen extrem de

[1] *Cele mai cunoscute sunt: Dicționarul limbii române (DLR, 2006), cunoscut drept "Dicționarul Academiei" (DA), accesibil mai ales specialiștilor; Mic dicționar academic, Ed. Univers Enciclopedic, 2001, 2002, 2003; Ion Coteanu, Luiza Seche, Mircea Seche (conducătorii lucrării) et al. (1975, 1996). Dicționarul explicativ al limbii române (DEX), București urmat de un supliment și alte versiuni, acum cea mai importantă resursă lingvistică online – www.dexonline.ro)*

[2] *Lanțul de preprocesări prin care este trecut corpusul nostru include un atribut, NotInDict, care semnifică faptul că nu a fost recunoscut de DEX-online (www.dexonline.ro), cea mai cunoscută resursă lexicală pentru limba română care servește scopurilor cercetărilor din domeniul procesării limbajului natural.*

Arhipelag XXI Press, Tîrgu Mureş, ISBN: 978-606-8624-03-7

BDD-V3969 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 00:54:25 UTC)

complex, aspectele diacronice sunt analizate pe baza unui model de calcul, la baza căruia se află un corpus de texte (aici, limbaj publicistic), din care sunt extrase caracteristici morfo-sintactice și semantice.

3. Resursele lexicale românești

Această secțiune este dedicată descrierii pe scurt a celor două resurse lexicale care au stat la baza acestui studiu. Este vorba de ***corpusul format dintr-o colecție de publicații*** din a doua decadă a secolului al 19-lea până în prezent corespunzătoare celor 4 regiuni analizate (v. Tabelul 1) și de Dictionarul Tezaur al Limbii Române în versiune electronică (eDTLR) (Fig. 1). Acesta din urmă servește scopului acestei lucrări, acela de a recupera cronologia unor cuvinte vechi extrase din colecția de ziare pe baza anului documentelor în care se regăsesc acestea.

De asemenea în tabelul 2, pentru exemplificare, am luat în considerare alte două tipuri de date, numele publicației monitorizate și regiunea unde a apărut, date care fac obiectul unor cercetări mai complexe în vederea măsurării statistice a similitudinii între limbi, cercetare publicată în diverse etape în (Dascălu & Gîfu, 2015; Gîfu & Simionescu, 2016; Gîfu, 2016).

Tabelul 1. Statistica generală a corpusului

A.	B.	C.	D	E.	F.
Regiunea	**Perioada**	**Total de cuvinte**[3]	**Totalul de cuvinte vechi**	**Totalul ocurențelor vechi**	**%(D/C)**
Moldova	1829-2015	65901	5085	2979	7.72
Țara Românească	1829-2015	137261	6525	4105	4.75

[3] *Din totalul de tokeni am exclus semnele de punctuație, numele și cuvintele mai mici sau egale cu 2 caractere.*

Arhipelag XXI Press, Tîrgu Mureș, ISBN: 978-606-8624-03-7

BDD-V3969 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 00:54:25 UTC)

Transilvania	1847-2015	160923	21023	8518	13.06
Basarabia	1848-2015	107324	4703	2891	4.38

Coloana A conține una dintre cele 4 regiuni geografice la care ne raportăm analiza, urmată de coloana B care indică perioada de monitorizare. Coloana C include numărul de cuvinte din care am reținut totalul de cuvinte vechi în coloana D. În coloana E apar doar ocurențele acestora. Coloana F ne arată ponderea cuvintelor vechi pentru fiecare colecție de texte.

Tabelul 2: Statistica parțială per regiune și publicație pentru perioada 1829-1848

1	2	3	4	5	6	7
Publicație	**Regiune**	**An**	**Total NotInDict**	**NotInDict unicat**	**Total cuvinte**	**% (4/6)**
Curier românesc	Wallachia	1829	11	11	207	5.31
Albina românească	Moldavia	1829	51	51	307	16.61
Albina românească	Moldavia	1832	20	20	152	13.15
Albina românească	Moldavia	1844	41	34	410	10
Organulu Luminarei	Transylvania	1847	769	610	1994	38.56
Cuvânt Dimitrie Balica	Bessarabia	1848	241	206	695	34.67
Organulu Luminarei	Transylvania	1848	393	282	974	40.34

E-DTLR este poate cea mai importantă resursă lexicală pentru informatizarea limbii române este *DTLR* și cuprinde o impresionantă bază de izvoare textuale, deocamdată accesibilă doar specialiștilor. [Cristea et al., 2007]. Se fac eforturi pentru crearea unei interfețe de redactare cu scopul de a transpune complexa arhitectură a dicționarului în categoriile universale ale standardului pentru reprezentarea textelor în

Arhipelag XXI Press, Tîrgu Mureș, ISBN: 978-606-8624-03-7

BDD-V3969 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 00:54:25 UTC)

format digital, TEI (*Text Encoding Initiative*), astfel încât dicționarul să aibă o structură similară și compatibilă cu dicționare ale altor limbi, descrisă mai jos.

articol

cuvânt titlu

elemente de calificare

sensuri nr

def

cit

informații privind pronunțarea

informații morfologice

variante grafice

variante lexicale

etimologie

DTLR are o structură semantică arborescentă, asemănătoare cu schema unui text discursiv. Este foarte importantă conceperea sensurilor cuvintelor ca pe niște relații de incluziune a subsensurilor și a unităților mai mici, sensuri figurate, expresii, sintagme, locuțiuni.



Figura 1: Interfața eDTLR

De exemplu pentru un verb, arborele conține 3 sensuri care corespund nivelului 3, unul dintre ele având un sub-sens corespunzător nivelului 5. (Curteanu et al., 2008:55).

<entry>

<hw>VERB</hw>

<senses>

<marker level=”3”>1.

Arhipelag XXI Press, Tîrgu Mureș, ISBN: 978-606-8624-03-7

BDD-V3969 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 00:54:25 UTC)

<definition>…</definition>
<marker level=”5”>◊
<definition>…</definition>
</marker>
</marker>
<marker level=”3”>2.
<definition>…</definition>
</marker>
<marker level=”3”>3.
<definition>…</definition>
</marker>
</senses>
</entry>

În dicționar se găsesc incluse, în aceste sensuri, unitățile frazeologice, dar și utilizări ale cuvântului care sunt marcate stilistic și nu sunt definite, ci ilustrate cu citate (folosite de noi pentru găsirea cuvântului NotInDict din colecția de ziare). De asemenea, aceste citate pot fi utilizate în alcătuirea unor corpusuri adnotate.

Ordinea sensurilor este una istorică, se pornește de la sensul cel mai vechi. Tot istorică este și ordonarea citatelor, care la fiecare sub-sens va fi selectat cel mai vechi citat. Acest lucru permite stabilirea unor filiații, urmărirea modificărilor diacronice suferite de sensul cuvintelor.

4. Metodologia de lucru

După procesarea textelor - care include segmentare, tokenizare, lematizare și analiza morfo-sintactică în care am inclus un atribut suplimentar pe care l-am notat cu NotInDict[4], utilizând POS-Tagger-ul de la Facultatea de Informatică a Universității „Alexandru Ioan Cuza” din Iași (Simionescu, 2011) – am extras aceste cuvinte cronologic pe care am început să le căutăm automat în eDTLR. (Fig.2).

[4] *Cuvintele care au această etichetă înseamnă că nu au fost recunocute de DEX-online (www.dexonline.ro), cea mai cunoscută și folosită în procesarea limbajului natural.*

Arhipelag XXI Press, Tîrgu Mureș, ISBN: 978-606-8624-03-7

BDD-V3969 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 00:54:25 UTC)

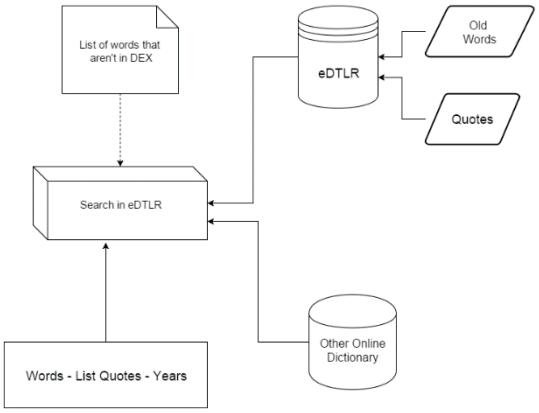


Figura 2: Arhitectura de recuperare cronologică prin citatele din eDTLR

Fiecare NotInDict este căutat în cuvântul-titlu care are asociat un set de referințe bibliografice, având atașat și anul publicării pe baza căruia vom decide dacă este înaintea celui pe care-l avem noi în corpus. Practic vorbim de două fișiere XML:

1. Primul conține citatele care includ și cuvântul nostru, marcat cu NotInDict. Acest prim fișier XML conține opera din care face parte acel citat adnotat prin marcajele <SG>...</SG>.

De exemplu: *dezmetit*

<entry accente="DEZMEȚIT" diacritice="DEZMETIT">

<list>**DEZMEȚÍT**, -Ă n-2</list>

<sense class="0" value="DEZMEȚÍT, -Ă ">

<definition><MorfDef>adj. </MorfDef><SpecDef>(Învechit)</SpecDef>

<RegDef> Dezmeticit (II 1). </RegDef>

<DefExemList>

<exemPair year="1857" year2="1858">

<DefExem>Trotușul e vechiul lor prieten, el îi primește în răcoroasele sale ape fără primejdii, îi îmbrățoșază și-i trimite pe acasă scăldați și **dezmețiți**. </DefExem>

<SG><AUTHCITE author="NEGRUZZI" pages="315" sigla="NEGRUZZI, S." source="S." volume="I">**NEGRUZZI, S. I, 315**</AUTHCITE>, cf. <AUTHCITE author="PONTBRIANT" sigla="PONTBRIANT, D." source="D.">PONTBRIANT, D.</AUTHCITE>, LM. **</SG>**

</exemPair>

<exemPair year="1881" year2="1882">

Arhipelag XXI Press, Tîrgu Mureș, ISBN: 978-606-8624-03-7

BDD-V3969 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 00:54:25 UTC)

<DefExem>Dar azi vălul cade, crudo! **dizmeţit** din visuri sece, Fruntea mea este trezită de al buzei tale-ngheţ. </DefExem>

<SG><AUTHCITE author="EMINESCU" pages="30" sigla="EMINESCU, O." source="O." volume="I">EMINESCU, O. I, 30</AUTHCITE>, cf. <SRCCITE source="DDRF">DDRF</SRCCITE>, RESMERIŢĂ, D.</SG>

<norm> - Scris şi: </norm>

</exemPair>**desmeţit**. - Pl.: <error>[i]dezmeţiţi, -te. - [/i]Şi:</error>

</DefExemList>

...

</sense>

</entry>

Aici am găsit mai multe citate care conţin cuvântul *dezmeţit*. Aşa cum spuneam ceea ce ajută la identificare automată a anului din care face parte un anumit citate este tagul <SG>.

Informaţiile pe care le va extrage după parsarea acestui fişier va fi stocat în unele structuri personalizate, după care se trece la următorul pas de procesare: analiza celui de-al doilea fişier XML din care se extrage anul asociat unei părţi sau întregului conţinut din sigla respectivă.

2. Al doilea fişier XML conţine date bibliografice eDTLR: mai multe elemente cu detalii privind un anumit autor/publicaţie, loc de publicaţie etc. Anterior am găsit ca simbol eticheta: “I. NEGRUZZI, S. I, 315“

Vom căuta unele elemente din eticheta precedentă (dat fiind că siglele nu sunt întotdeauna scrise unitar) şi vom extrage valoarea atributului care ne interesează; anul de apariţie.

<RandSigla nr_item="2686">

<Cronologie **an**="1889">168</Cronologie>

<Sigla>I. NEGRUZZI, S. I–VI</Sigla>

<Autor> Iacob Negruzzi</Autor>

<Titlu>Scrieri complecte.</Titlu>

<Detalii> Vol. I–VI. I: Copii de pe natură. Scrisori; II: Poesii; III: Mihai Vereanu. Primblări prin munţi; IV: Teatru; V: Pe malul mării. Din Carpaţi. Traduceri din Schiller; VI: Traduceri din Schiller. Bucureşti, Socec & Comp., 1893–1897.</Detalii>

Arhipelag XXI Press, Tîrgu Mureş, ISBN: 978-606-8624-03-7

BDD-V3969 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 00:54:25 UTC)

<auxiliar> Iacob Negruzzi, Scrieri complecte. Vol. I–VI. I: Copii de pe natură. Scrisori; II: Poesii; III: Mihai Vereanu. Primblări prin munţi; IV: Teatru; V: Pe malul mării. Din Carpaţi. Traduceri din Schiller; VI: Traduceri din Schiller. Bucureşti, Socec & Comp., 1893–1897.</auxiliar>

<Pagini>838</Pagini>

<NrPaginiVolum>I-492, II-346</NrPaginiVolum>

<locatie>IFRI</locatie>

<VecheRara/>

<COPYRIGHT>NU</COPYRIGHT>

<stiute>1</stiute>

<copyright>0</copyright>

<slavona>0</slavona>

<SCANATE>scanat doar I si II, restul scanat BCU</SCANATE>

<scanate>1</scanate>

</RandSigla>

Pentru o și mai clară exemplificare luăm un cuvânt din ziarul Pressa, 3 august 1878, București, după cum urmează:

- **Input inițial – forma din text**:

PRESSA, BUCURESCI, 3 AUGUST 1878

[..] *Dar toate aceste modificări de* ***Constituţiune****, adică recunoscerea şi aplicarea principielor puse în* ***tractatul*** *din Berlin, se pot face altfel, decât prin o Constituantă! La aceste modeste* ***cuestiuni*** *rugăm să ne răspundă, limpede şi categoric,* ***ghibacii****, isteţii şi mult naţionalii redactori de la Românul.*

- **Input inițial – extragem lista de cuvinte** marcate cu NotInDict.

[Constituţiune, tractatul, cuestiuni, ghibacii]

- **Ca output:**

- **Căutăm cuvântul după definiție și citate:**

Singurul cuvânt din lista de mai sus pe care l-am găsit în eDTLR este *ghibac*.

<entry accente="**GHIBACIU**" diacritice="GHIBACIU">

<list>GHIBÁCIU, -CE || || n-4</list>

Arhipelag XXI Press, Tîrgu Mureş, ISBN: 978-606-8624-03-7

BDD-V3969 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 00:54:25 UTC)

<sense class="0" value="GHIBÁCIU, -CE">

<definition>

<MorfDef>adj., adv. </MorfDef>

<ItalMarker>Adroit. Adroitement. -

<norm>

Îndemânatec și isteț (la lucru), iscusit, meșter, abil, cu apucătură. A d j.</norm>Omeșteșugie ghibace, de care ne putem cu dreptul mirà.<norm>

<SG>RUSSO, S. 60.</SG>

</norm>

Babele... sânt foarte dibace în tragerea bobilor.

<norm>

<SG>ALECSANDRI, P. P. 234.</SG>

[…]

3. Căutăm cuvintele din lista anterioară după alte citate decât cele în care a fost definit:

i. *După ce Petru încheie un* ***tractat*** *demărcătoriu ... între China şi Rosia ... socoti a-şi organiza o armie regu-lată.* - SĂULESCU, HR. II, 456/19

ii. *Bogdan însă dizvălea o politică ghibace*. - ASACHI, S. L. II, 24

4. Căutăm sigla cu autorul/autorii din care a/au fost găsit citatul/citatele în care am găsit cuvântul din lista de mai sus:

5.

SĂULESCU, HR. II, 456/19 - 1834

<RandSigla nr_item="3124">

<Cronologie an="1834">93</Cronologie>

<Sigla>SĂULESCU, GRAM. ROM. I–III</Sigla>

<Autor>Gh. Săulescu</Autor>

<Titlu>Gramatică românească sau observaţii grămăticeşti asupra limbei româneşti. Pentru şcoalele normale şi ghimnaziale. Partea I-ia etimologhică, II sintactică, III de

Arhipelag XXI Press, Tîrgu Mureş, ISBN: 978-606-8624-03-7

BDD-V3969 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 00:54:25 UTC)

prosodia limbei româneşti şi de versificaţie, la care s-au adaos mai multe exempluri de versuri.</Titlu>

<Detalii>Iaşi, Tipografia Sfintei Mitropolii, partea I–II: 1833; partea a III-a: 1834.</Detalii>

<auxiliar>Gh. Săulescu, Gramatică românească sau observaţii grămăticeşti asupra limbei româneşti. Pentru şcoalele normale şi ghimnaziale. Partea I-ia etimologhică, II sintactică, III de prosodia limbei româneşti şi de versificaţie, la care s-au adaos mai multe exempluri de versuri. Iaşi, Tipografia Sfintei Mitropolii, partea I–II: 1833; partea a III-a: 1834.</auxiliar>

<Pagini>234</Pagini>

<NrPaginiVolum>I – 234</NrPaginiVolum>

<locatie>IFRI</locatie>

<VecheRara>CV</VecheRara>

<COPYRIGHT>NU</COPYRIGHT>

<stiute>1</stiute>

<copyright>0</copyright>

<slavona>1</slavona>

<SCANATE>scanat</SCANATE>

<scanate>1</scanate>

</RandSigla>

6. ASACHI, S. L. II, 24 - 1838

<RandSigla nr_item="340">

<Cronologie an="1838" an2="1839">99</Cronologie>

<Sigla>ASACHI, S. L. I–II</Sigla>

<Autor>Gh. Asachi</Autor>

<Titlu>Scrieri literare.</Titlu>

<Detalii> [Vol. I şi II]. Ediţie îngrijită, cu prefaţă, note şi glosar de N. A. Ursu. [Bucureşti], Editura de Stat pentru Literatură şi Artă, [1957]. (Biblioteca pentru toţi).</Detalii>

<auxiliar>Gh. Asachi, Scrieri literare. [Vol. I şi II]. Ediţie îngrijită, cu prefaţă, note şi glosar de N. A. Ursu. [Bucureşti], Editura de Stat pentru Literatură şi Artă, [1957]. (Biblioteca pentru toţi).</auxiliar>

Arhipelag XXI Press, Tîrgu Mureş, ISBN: 978-606-8624-03-7

BDD-V3969 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 00:54:25 UTC)

<Pagini>654</Pagini>
<NrPaginiVolum>I – 286; II – 368</NrPaginiVolum>
<locatie>IFRI</locatie>
<VecheRara>–</VecheRara>
<COPYRIGHT>DA</COPYRIGHT>
<stiute>1</stiute>
<copyright>1</copyright>
<slavona>0</slavona>
<SCANATE>scanat</SCANATE>
<scanate>1</scanate>
</RandSigla>

Statistic pe baza acestei metode am descoperit faptul că mai bine de jumătate din aceste cuvinte, au intrat în limbă cu un secol, chiar două mai repede.

Pentru ilustrare, Tabelul 3 conține câteva exemple.

Tabelul 3: Recuperarea cronologică a cuvintelor vechi din colecția de ziare 1829-2015 (mostră)

Corpus publicații 1929-2015				**eDTLR**		
NotInDict	**An**	**Regiune**	**Publicația**	**An**	**Citat**	**Date sursă**
abie	1868	Transilvania	Federațiunea	1750	Un om… căzut-au în tălhari de-l jăfuire și-l răniră și-l lăsară abie viu în cali	M. Gaster, Chrestomaţie română. Texte tipărite şi manuscrise (sec. XVI–XIX), dialectale şi populare, cu o introducere, gramatică şi un glosar româno-francez de… Vol. I: Introducere, gramatică, texte (1550–1710).
acestia	1854	Transilvania	Foaia pentru Minte, Anima și Literatura	1704	Vredniciile acestia să cuvin lui Petru, de vreme ce el... au întins pururea, ca alt soare, de taină, peste tot pământul	Studii şi cercetări lingvistice. [Bucureşti], Editura Academiei. Institutul de Lingvistică din Bucureşti. Anul I (1950) ş.u.

Arhipelag XXI Press, Tîrgu Mureş, ISBN: 978-606-8624-03-7

BDD-V3969 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 00:54:25 UTC)

					razele darurilor și a facerilor de bine.	
semni	1950	Transilvania	Vatra românească	1773	Pi semni, omule, tu eşti turuielic	M. Gaster, Chrestomaţie română. Texte tipărite şi manuscrise (sec. XVI–XIX), dialectale şi populare, cu o introducere, gramatică şi un glosar româno-francez de… Vol. I: Introducere, gramatică, texte (1550–1710).
simtu	1853	Transilvania	Telegrafulu Romanu	1600	Cum tămăduieşti pre alţii, ce simtu mai streini, aşea vino de tămăduiaşte şi pre prietinul tău, pre Lazăr	M. Gaster, Chrestomaţie română. Texte tipărite şi manuscrise (sec. XVI–XIX), dialectale şi populare, cu o introducere, gramatică şi un glosar româno-francez de… Vol. I: Introducere, gramatică, texte (1550–1710).
sîngur	1950	Transilvania	Vatra românească	1762	Vel-cupar, cu tipsia cea de argint, stînd înaintea domnului, pune vutcă în doao păhare şi întîi cuparul ia credinţă, apoi domnul sîngur ia paharul de-l pune la gură şi ia şi confeturi.	Dan Simonescu, Literatura românească de ceremonial. Condica lui Gheorgachi, 1762. Studiu şi text de... Bucureşti, Fundaţia Regală pentru Literatură şi Artă, 1939. [Secţia istorică, III. Studii].
soarile	1848	Basarabia	Candela	1750	La Eghipet nu cade niciodată zăpadă, pentru că este soarele mai aproape şi în veac	M. Gaster, Chrestomaţie română. Texte tipărite şi manuscrise (sec. XVI–XIX), dialectale şi populare, cu o introducere, gramatică şi un

Arhipelag XXI Press, Tîrgu Mureş, ISBN: 978-606-8624-03-7

BDD-V3969 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 00:54:25 UTC)

					soarile și văzduhul este mai herbinte, precum vestesc cei ce lăcuiesc acolo	glosar româno-francez de… Vol. I: Introducere, gramatică, texte (1550–1710).
svârli	1950	Transilvania	Vatra românească	1688	Și mâncă Iacov și să sătură și svârli cel iubit	M. Gaster, Chrestomaţie română. Texte tipărite şi manuscrise (sec. XVI–XIX), dialectale şi populare, cu o introducere, gramatică şi un glosar româno-francez de… Vol. I: Introducere, gramatică, texte (1550–1710).

5. Concluzii și direcții viitoare de cercetare

Așa cum era de așteptat, semnificația resurselor lexicale prezentate (eDTLR și colecția de publicații din România și Basarabia) pentru lingvistica computațională românească este de necontestat. În principal, eDTLR ușurează semnificativ efortul categoriilor de utilizatori care doresc fie să se informeze asupra sensului unui cuvânt (mai ales pentru formele vechi ale cuvintelor), fie să întreprindă cercetări complexe asupra limbii române, cum este și un asemenea studiu.

Din cele relatate, eDTLR având în componența sa o varietate impresionantă de documente vechi, răspunde foarte bine oricărei cercetări care vizează recuperarea crononologică a unor cuvinte identificate în texte din secolul al 19-lea în prezent.

Dezvoltarea celor două resurse lexicale este un proiect în continuă dezvoltare, ele putând fi folosite în mod coerent, dar transparent pentru o multitudine de aplicații lingvistice (analize sintactice, analize semantice, dezambiguizare de sens etc.).

Mulțumiri: Această lucrare a fost publicată cu sprijinul sprijinul studneților din anul al II-lea de la masteratul de Lingvistică computațională, de la Facultatea de Informatică a Universității “Alexandru Ioan Cuza” din Iași.

Referinţe bibliografice

Arhipelag XXI Press, Tîrgu Mureş, ISBN: 978-606-8624-03-7

BDD-V3969 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 00:54:25 UTC)

1. Avram, M., Sala, M., Vintila-Rădulescu, I. (coords.): *Institutul de Lingvistică „Iorgu Iordan". 50 de ani de existență (1949-1999)*, Univers Enciclopedic, București, 1999.

2. Coteanu, I., Seche, L., Seche, M. (coords.) et al.: *Dicționarul explicativ al limbii române* (DEX), București, 1975, 1996.

3. Curteanu, N, Moruz, A., Trandabăț, D.: *Extracting Sense Trees from the Romanian Thesaurus by Sense Segmentation & Dependency Parsing*. In: Proceedings of the Workshop on Cognitive Aspects of the Lexicon (COGALEX 2008), Manchester, 2008, pp. 55.

4. Danzin, A.: *Towards a European Language Infrastructure*, March, 1992.

5. *Dicționarul Limbii Române*, Tomul I, Partea a 4-a, Litera D, Deja-deținere, București, Editura Academiei Române, 2006.

6. Dascălu, M. and Gîfu, D.: *Evaluating The Complexity Of Online Romanian Press*. In: Proceedings of the 11th International Conference Linguistic Resources and Tools for Processing The Romanian Language, ConsILR-2015, 26-27 Nov. 2015, Iași, Romania, pp. 149-162.

7. Gîfu, D.: *Diachronic Analysis Using a Statistical Model* at the Mathematical Foundations of Informatics Conference, MFOI 2016, second edition, 25-29 July 2015, Chișinău, Republic of Moldova.

8. Gîfu, D. and Simionescu, R.: *Tracing Language Variation for Romanian* at the 17th International Conference on Intelligent Text Processing and Computational Linguistics, CICLing 2016, 3-9 Apr. 2016, Konya, Turkey.

9. *Mic dicționar academic,* Ed. Univers Enciclopedic, 2001, 2002, 2003.

10. Tufiş, D., Slama-Cazacu, T.: *An Overview of Research and Development in Language Technology in Romania*. In: Proceedings of Conference Awareness Days on Language & Technology, Luxembourg, 1994.

11. Vintilă-Rădulescu, I.: *Resurse lingvistice pentru limba română elaborate la Institutul de Lingvistică „Iorgu Iordan"*. In: Limba Română în Societatea Informațională – Societatea Cunoașterii, Tufiș, D., Filip, F. Gh. (coord.), Ed. Expert, București, 2002, pp. 21-33.

12. Simionescu, R.: *UAIC Romanian Part of Speech Tagger*, resource on nlptools.info.uaic.ro, "Alexandru Ioan Cuza" University of Iași, 2011.

Arhipelag XXI Press, Tîrgu Mureș, ISBN: 978-606-8624-03-7

BDD-V3969 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 00:54:25 UTC)

Webografie

www.dexonline.ro

Arhipelag XXI Press, Tîrgu Mureş, ISBN: 978-606-8624-03-7

BDD-V3969 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 00:54:25 UTC)